INTEGRATING HETEROGENEOUS DATA REPRESENTATIONS IN MODEL-BASED AEC/FM SYSTEMS

Branka Kosovac[1], Thomas M. Froese[2], and Dana J. Vanier[3]
1 Department of Civil Engineering, University of British Columbia
2 Department of Civil Engineering, University of British Columbia
3 Institute for Research in Construction, National Research Council Canada

ABSTRACT: The paper introduces an approach to the management of heterogeneous representations of information in integrated AEC/FM systems. The approach is based on the concept of controlled vocabularies — thesauri. In this context, a thesaurus is used to control not only terminology but all different representations or displays of information, with a focus on non-textual representations. Examples of different representations include: graphical representations at varying level of detail, diverse graphic formats (vector, pixel, VR, video etc), graphics-text-audio as both alternative and complementary modes, visualized non-natively-graphic information etc. The approach is illustrated by a set of simple applications, the first in a series of interrelated prototypes comprising an intended research project. The applications use HTML, JavaScript and XML along with an existing pilot thesaurus in the low-slope roofing domain for managing (i.e. storing, relating, and retrieving) textual and non-textual information from both local and remote sources. The applications represent a proof of the concept, to be elaborated in future work, that will used for efficient and economical management of information in construction planning and facilities management applications within integrated model-based systems.

KEYWORDS: Information management, thesauri, integrated model-based systems, XML, heterogeneous representations of information

1. INTRODUCTION

Two ongoing, recently integrated, industry-wide initiatives in the AEC/FM domain—the Industry Foundation Classes (IFC) (IAI 1999) and aecXML (aecXML.org 1999), share both a common goal and the same basic means. The goal is interoperability and information re-use within the domain; while the means is the enrichment of data sets by machine-readable domain semantics. In this paper we point to some problems that still remain with the implementation of these standard schemas in practice and we suggest the use of yet another semantics-based tool and related techniques for enhancing information utilization in the AEC/FM domain—the use of thesauri (Vanier, 1994, Kosovac et al., 2000).

1.1 Vocabulary problem

Domain schemas outline relationships between concepts pertaining to a certain domain. Concepts are traditionally, and usually most conveniently, represented by words, since words are the most widely accepted conventions for designating classes of “things”. However, different users and different applications may use a variety of terms for the same concept. Furnas et al. (1987), for example, found that the probability of two people using same terms for objects from five different domains was less than 20%. Humans can also easily relate the meanings of two terms as "the same", "similar", "more specific", "more general" and so on. On the other hand, computers can handle only strings, not concepts, and they are not able to relate different terms to the same referent without explicit instructions. The semantics of a domain model is machine-usable only if it is expressed using an agreed upon vocabulary and

syntax. The mismatch between vocabularies used within a system, known as the "vocabulary problem" in the field of information science, has strongly negative effects on the efficiency of information use. It can be expected that model-based applications will be used within a globally networked environment, which offers a wealth of information that is relevant to particular AEC/FM projects and potentially directly usable by the applications. Naturally evolving and driven by free market forces, this wider environment inevitably features heterogeneity. One aspect of the heterogeneity is the use of diverse vocabularies. Furthermore, the international scopes of the IFCs and the Internet imply the use of different languages and different variants of the same language. As all access to information is, in fact, based on string matching, it is quite clear that some form of vocabulary control might enhance information reuse within a domain by integrating information from sources using alternative vocabularies.

1.2 Heterogeneous representations of information

The same problems can be observed in the more general terms of information representations, which may vary in format (e.g. descriptive text, database table, database report, STEP physical file, video, audio etc.), scope (e.g. building, basement, mechanical system), and level of detail (schematic vs. photorealistic vs. shop drawing, general vs. subcontractor's schedule). A photograph, a CAD model, and a free-hand sketch can represent the same building; a particular cost information can describe a more general class of products for a certain design information; a specific contract documents relate to parts of particular schedule activities, and so on. Such facts may be obvious to humans but computers again need instructions.

2. PROPOSED APPROACH

We suggest that semantic interoperability needs integrated solutions at three levels—the levels of: (1) concepts, (2) machine-readable/understandable representations or data structures, and (3) human-understandable presentations or displays. The ongoing integration efforts have only been focusing on the first and second level of interoperability, that is, on the development of conceptual models and their translation into machine-readable form. We are addressing issues related to human-understandable representations of information and relation thereof to the remainder of the mechanism, in belief that solutions at this level can enhance and extend the reach of information integration in the domain. As a remedy to the above outlined problems, we propose the use of thesauri—structured controlled vocabularies originally devised as a solution to the vocabulary problem in information storage and retrieval systems.

2.1 Thesauri

In addition to ontological relationships between concepts, which are also identified in domain schemas, thesauri identify lexical or terminology relations between terms designating these concepts. In that way, thesauri span two levels of abstraction (Jarvelin et al., 1996):

- conceptual, including generic (broader term/wider term), partitive (whole term/part term), and associative (related term) relationships, and
- linguistic, encompassing delineation of terms' scope (or meaning), synonyms linking through the equivalency relationship (USE-USED FOR, USE ... AND), and disambiguation of homographs (e.g. bridges[structures] vs. bridges[dentistry]) (National Information Standards Organization, 1993).

The frequently overlooked distinction between concepts and terms can be easily translated into the distinction between content and form, signifier and signified, or information and information representation. From this broad perspective, it can be seen that thesaurus mechanisms can be applied much wider—to integrated management of information at two levels of abstraction: conceptual and representational. In other words, we suggest that

bringing together alternative representations of objects or concepts on the one hand and semantic networks encompassing their hierarchical and associative relationships on the other, using some tokens (e.g. "preferred terms" or IDs) can enhance information management in AEC/FM systems. The meaning of the term system in this context is ultimately broad—it means both the totality of potentially networked information in the domain and, at the same time, any independently functional component of an application.

2.2 Categories of information

The heterogeneity of the wider information environment also relates to the coexistence of structured and unstructured information—two basic categories, each encompassing a whole range of variations. We can observe differences between these categories in terms of their conceptual schemas and their inherent machine and human understandable representations.

Structured information implies machine understandable representations highly structured according to a certain conceptual schema. A variety of human understandable displays can be derived from a single machine representation: a schematic 2D drawing, a rendered 3D model, or a process simulation, for example. This category includes, first, data highly structured according to IFCs, aecXML, or another comprehensive and detailed domain schema. However, even if a model is accepted as an industry-wide standard, there will always be sub-domains, related domains, and applications that are better served by schemas with partly overlapping scopes and structures of their own. Important examples of alternative schemas are the widely used classification systems, such as CSI in the United States or BSAB in Sweden. Another important and very broad subcategory includes information that may be highly structured, but structured according to a fundamentally different schema—typically based on information display and not on information content. For convenience, we will treat this category separately and call it, although possibly inadequately—*ill-structured information*. It includes a variety of primarily textual information such as building codes, web pages, word processor documents, or user queries, which are structured according to a document model (data type definition or DTD), HTML specification, or a specific query syntax. The same category also includes primarily visual information such as CAD databases whose schemas involve objects such as lines and shapes and not doors, columns and roofs.

Unstructured information encompasses unstructured or barely structured text (such as most word processor and HTML documents) and a variety of non-textual information such as bitmap images, video, audio etc. The content of such representations is understandable to humans but not to machines, which can read and handle such data only at the low level of pixels or simple strings or as complete, fixed documents and display it in a single, fixed way.

2.3 Possible uses of thesauri

The original purpose of thesauri was to assist in the selection of indexing terms (keywords) for journal articles at both ends of information storage and retrieval systems. This use has been significantly extended over time (Kosovac, 1998), especially to indexing a wide range of materials and to various uses in natural language processing (NLP) applications. We believe that thesauri and thesaurus principles may be used with any of the aforementioned categories of information.

In the case of highly structured information, the addition of lexical relations to the semantic relationships already defined by the schema might extend the usability of model-based systems—user interfaces and access to external information in particular. An important issue in model-based systems involves mappings between the core models and view models, and between alternative domain schemas. Despite multiple research efforts in the area of inter-

schema mapping (Verhoef et al., 1995), and the ongoing development of the standard mapping language (i.e., Express-X) (ISO, 1999), mapping problems are still far from being solved (Eastman, 1999). Thesaurus services might be used in the development of schema mappings or for "on-the-fly" mappings, (e.g. for mapping terms used in queries or information requests originating from one model/system to terms used as labels and values in another).

In the case of unstructured or ill-structured textual information, in addition to their classical role for assigning and searching metadata, thesauri can be used for searching and retrieving free text or integrated into data mining, knowledge extraction, or other related applications. For example, building codes for the most part have strong typographical structures and use controlled vocabulary but could benefit from semantic classification for the creation, modification and retrieval of information (Vanier, 1994).

Similar uses of thesauri are applicable to unstructured non-textual sources. As mentioned earlier, thesauri are widely used for assigning metadata to different kinds of materials, such as photographs, videos, sound recordings, museum exhibits, etc. This use, if seen as mapping of human-understandable representations to concepts, can be extended to relating such representations (e.g. photographs of real buildings, paper drawings, video-clips of construction processes, unstructured text, symbols etc.) to structured data. The mechanism provided by the IFCs is one-way referencing of a document (i.e. computer file defined by a set of document management-related properties within IfcDocumentResource) from a model (IfcObject) through the objectified IfcDocumentReference relationship. Linking non-textual materials to model objects involves some intricate issues extensively investigated in the field of indexing and retrieving pictorial materials. Fortunately, there are some mitigating circumstances in the AEC/FM domain. One is that, unlike in many other domains, pictorial materials (e.g. sketches, 3D models), the primary type of non-textual representations in the domain, are to a great extent denotative. This means that complex and unavoidably subjective analysis of connotative meanings, as needed, for example, in fine arts or general picture collections, and which represents the most challenging issue in managing images, needs not to be considered. Another mitigating circumstance is that the users and possible uses of such materials are reasonably predictable. The problem that still remains is the level of specificity and level of detail at which these documents should be related to model objects. Computer vision, automatic feature extraction, image recognition, and other related techniques represent a counterpart to free text searching and data mining. It is important to note, however, that all the up-to-date advancements in this field do not diminish the significance of the concept-representation mappings, since mechanisms for aggregation and abstraction of extracted features and "linking low-level visual features with high-level semantics" (Chang et al., 1997) are needed.

Interlacing all these discussions is the notion that thesauri can serve for linking all different types of information and their representations in the domain.

3. IMPLEMENTATION ISSUES

3.1 Model of data exchange

We can observe an analogy between the process of data exchange and that of information storage and retrieval. Buckland and Plaunt (1994) developed a "minimally complete model" outlining "the functional components that appear to be necessary and sufficient to represent information storage and retrieval systems." Figure 1 shows a diagram of the model as we applied and adapted it to data exchange. It is important to note that real systems do not

necessarily include all of the components; the only necessary process is "matching" or, according to Buckland's and Plaunt's terminology—"partitioning". Systems may include multiple components of the same type, system boundaries are arbitrary, and there may be feedback loops in any of the component processes. As such, this model can be applied to a variety of environments, from components of standalone applications to large distributed systems managing project information or company archives to external libraries of products or construction methods, all the way to the overall flow of information in the domain. Another important point is that each of the processes in the model can be elaborated into a separate system structured again according to the same model and acting as a module that interacts with the rest of the larger system.

External knowledge sources include domain schemas, thesauri, classification systems etc. As shown in figure 1, they can be used for multiple purposes: for creating system's internal representations of source objects, for creating searchable indices, for translating user query/information request into a formal query, and for translating retrieved data into a suitable output. Ideally, from the data-exchange point of view, the same external knowledge source would be used throughout a system, for all of the listed purposes. That is the idea underlying the use of standard data models. However, as noted earlier, this is not always achievable.

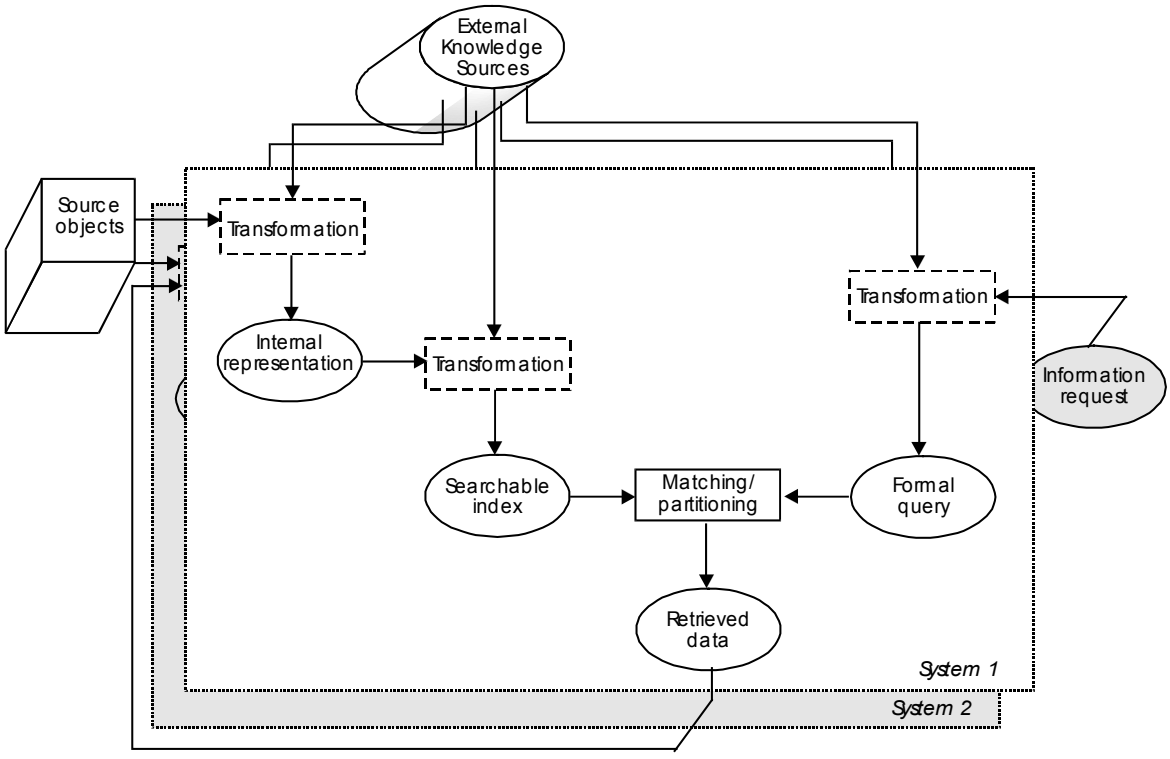

Figure 1. Model of data exchange

3.2 Possible levels of control/scope

The aforementioned scalability of the model implies that it is possible to implement thesauri at different levels of control, from sub-domain, region or company-specific to industry-wide. One problem is that the increasing scope of a thesaurus enhances its potential acceptance/use, and thus directly its usefulness, and on the other hand, encumbers its development, maintenance and the flexibility necessary to accommodate the inevitable variety of needs. An increasingly popular choice is to provide mappings between heterogeneous vocabularies

and classification systems used in a domain. The National Library of Medicine's (1999) Unified Medical Language System (UMLS), which enables efficient retrieval and integration of biomedical information, represents an outstanding example of centralized control of existing heterogeneous sources. This illustrious effort is extremely labour-intensive, as it relies on human expertise. Another example is the ongoing CERES/BRD Thesaurus Partnership Project, where a core environmental thesaurus is being developed and linked to the existing controlled vocabularies along with templates for the development of state-specific thesauri. (CERES, 1998). Automatic inter-thesaurus mapping has been a frequent topic of research and development (e.g. Amba et al., 1996; Buckland et al., 1999) but none of the reported successful approaches has become widely accepted or used as an exclusive means for inter-thesaurus mapping in a domain yet.

The choice of the control level and mechanisms will certainly depend on the nature of the domain (i.e. heterogeneity of participants, patterns and complexity of information flow, information formats etc.) and on the existing infrastructure (i.e. existing technology, schemas, vocabularies, standards, associations and regulating bodies etc.).

3.3 Implementing a thesaurus service

The ruling trends in information technology (e.g. modularity, extensibility, flexibility, Internet orientation), reflected in all ongoing standardization efforts, coupled with the characteristics of information flow in the AEC/FM domain (i.e. complex information exchanges between a large number of diverse participants), suggest distributed systems as the prevailing future environment. The overall architecture for distributed information systems can be thought of in terms of the software components that make up the system and the services—or packages of functionality—offered by these components (Froese and Yu, 2000). Much of the thesaurus capabilities discussed in this paper could be modularised into an independent component that offers thesaurus services. This component would typically be implemented as a server-based component accessible over the Internet through a standard application programming interface (API), although it could also reside on local workstations for certain applications.

A thesaurus services component would contain an underlying “thesaurus engine” featuring capabilities of thesaurus management software. Within the AEC/FM domain, a number of thesaurus vocabularies covering any distinct but easily manageable sub-domain would be constructed as the need arises. The vocabularies would be made publicly available in a repository and based on the existing schemas wherever possible. The adherence to existing standards (e.g. ANSI/NISO 39.19, thesaurus XML schemas, Z39.50) is strongly desirable as well as mappings to a common core, most likely the existing IFCs or a portion thereof.

The thesaurus services component would have no direct user interface, but would make its services available to applications and other software components through its API. For example, one simple application could allow users to directly browse textual and graphical thesauri by implementing a user interface for the capabilities offered through the thesaurus services component. In other cases, different components of distributed AEC/FM information systems would make use of the thesaurus services component, such as components that offers query services on a model-based project database, schema mapping components, data warehousing components, etc.

4. PROTOTYPE THESAURUS MODULES

This research project comprises a series of interrelated prototypes investigating particular issues involved in the management of heterogeneous information representations. The first

set of prototypes uses a pilot thesaurus in the domain of low-slope roofing to test, at a small scale and at the basic level, different aspects of our expectation that the complexity of AEC/FM information can be managed by using networks of tokens representing concepts or objects, their relationships, and diverse representations.

The thesaurus was initially implemented as a set of hyperlinked HTML files allowing browsing of thesaurus terms for the purposes of familiarization with the domain or selecting keywords or search terms. The subsequent addition of HTML forms/buttons and JavaScript allowed searching unstructured information on the World Wide Web by sending simple queries (truncated phrases) to search services.

The next step involved the extension of the thesaurus's function to non-textual representations. This prototype, shown in figures 2 and 3, demonstrates the synergy of text and graphics as alternative/complementing accessing structures and display modes. Thesaurus terms can be accessed by browsing an alphabetical list of terms, displays of groups of interrelated terms or through pictorial representations of wider concepts. The prototype illustrates the ability to access and navigate AEC/FM information using diverse graphical representations, with all the linking done using "preferred terms" as tokens.

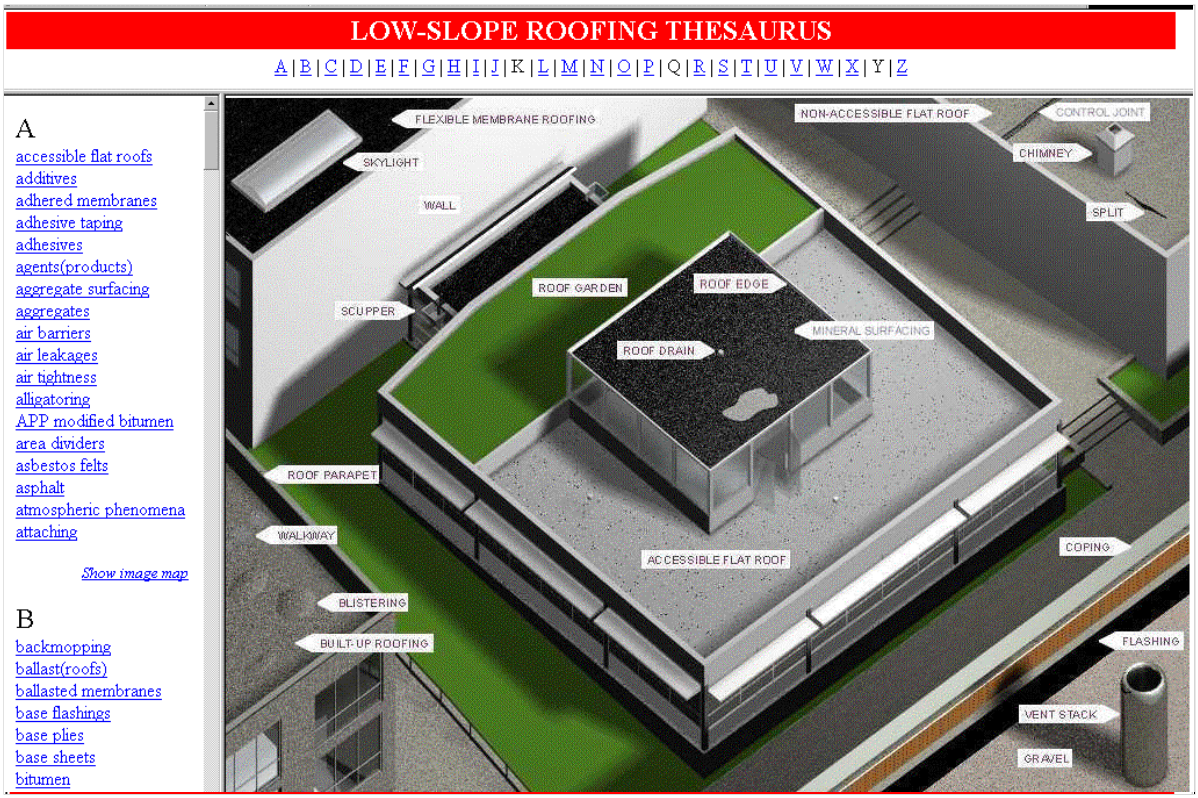

Figure 2. Screen shot of the HTML prototype, illustrating multiple accessing structures.

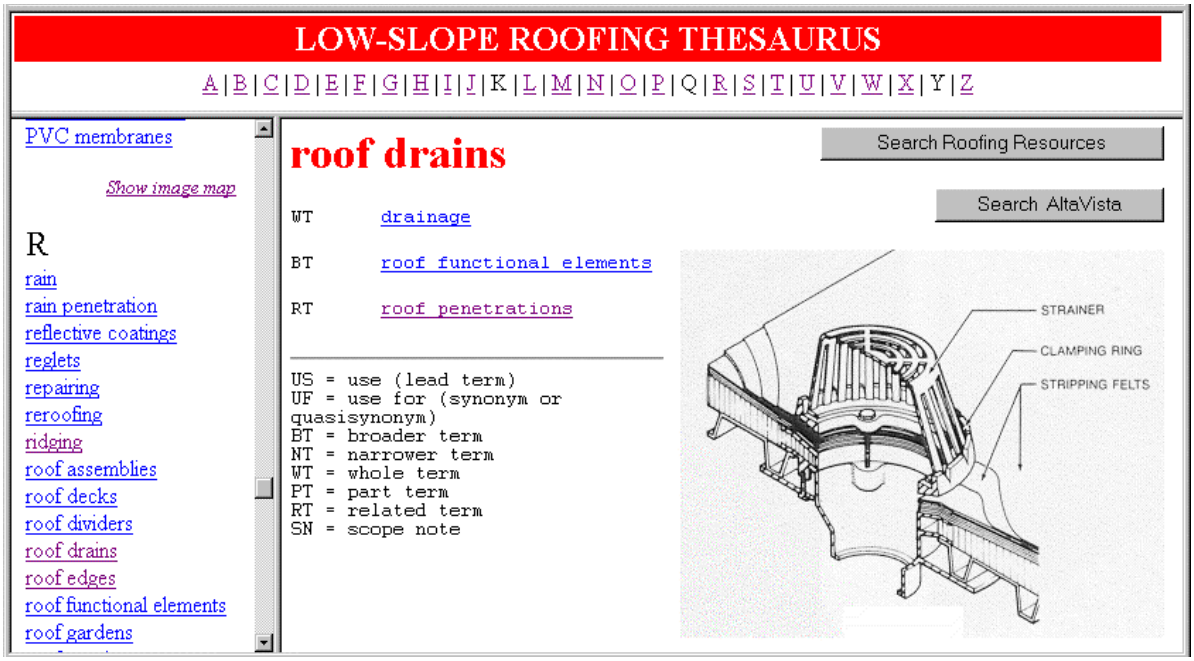

Figure 3. Screen shot of the HTML prototype, illustrating complementing displays.

Our latest prototype, shown in figure 4, investigates the potential of XML for enhancing the application. The pilot thesaurus was converted into an XML file. Although there are at least two general DTDs/XML schemas for thesauri proposed (Taylor, 1999 and Lee et al., 1999), a custom schema was developed and used which best befitted the structure of the existing thesaurus and the needs of the application. However, to maximize the potential of the tool, the use of standard schemas is suggested. JavaScript was used for accessing and manipulating the DOM tree nodes in order to provide the following functionality:

- Simple HTML display that allows browsing and selection of thesaurus terms
- Launching searches with automatic expansion of the selected term with its synonyms to one general (AltaVista) and one specialized (Roofing Resources) search service
- Expansion of AltaVista searches with narrower and part terms upon user's request
- Browsing a local collection of images

Although one obvious use of thesaurus modules would be searching/querying XML files—both XML tags and values, it was not tested in this project. One reason is that suitable corpora were not available and coding documents specifically for this purpose would give biased results, the other was the early stage of development of the standard XML Query language (W3C, 2000), which should be used for that purpose, rather than some custom script. A manual matching of the thesaurus to the only known domain-related XML file—the aecXML specification gave us some useful guidelines for the development of thesauri that need to be mapped or used in mappings. First, a strong emphasis should be put on the exhaustive identification of synonyms. The use of automation for collecting a wide range of candidate terms is advisable (e.g. as described in Kosovac et al., 2000). Second, compound terms should be used sparingly, relying on the function of *USE ...+...* cross-references. Third, it may be useful to organize thesaurus terms into broad categories—facets, as this might facilitate post-co-ordination, mapping to different classification schemes, and indexing of unstructured information.

The early stages of development and consequent difficulties in implementation also prevented the exploration of some other ancillary XML standards. XML Linking Language

(XLink) (W3C, 1999) has a potential to enhance applications of this type since XLinks allow linking to multiple resources, explicit mechanisms to associate metadata and special behaviour with the link, bi-directional linking, linking to databases and their special functions etc. Extensible Style Language (XSL) and, possibly, Scalable Vector Graphics (SVG) promise great flexibility of thesaurus displays.

For practical reasons, instead of assigning thesaurus terms to images, we included pointers to image files (URLs) into the nodes representing the corresponding indexing terms. Indexing/linking was done following a combination of guidelines from (Shatford 1986), the subject analysis scheme devised at the Bibliothéque National in Paris (Preaud and Rio 1978), Guide to the Description of Architectural Drawings (Foundation for Documents of Architecture, 1999) and Thesaurus for Graphic Materials (Library of Congress, 1999). The script traverses XML nodes and retrieves, not only images directly representing a concept/object, i.e. referred from its node, but also representations of concepts including that concept (through partitive relationships), abstractions and more specific examples (through generic relationships) etc. where applicable and requested. This prototype shows how a well-developed network of concept IDs, which represents conceptual relationships between terms in the domain, can also bring together alternative representations of the concepts that can be used for more sophisticated access and navigation.

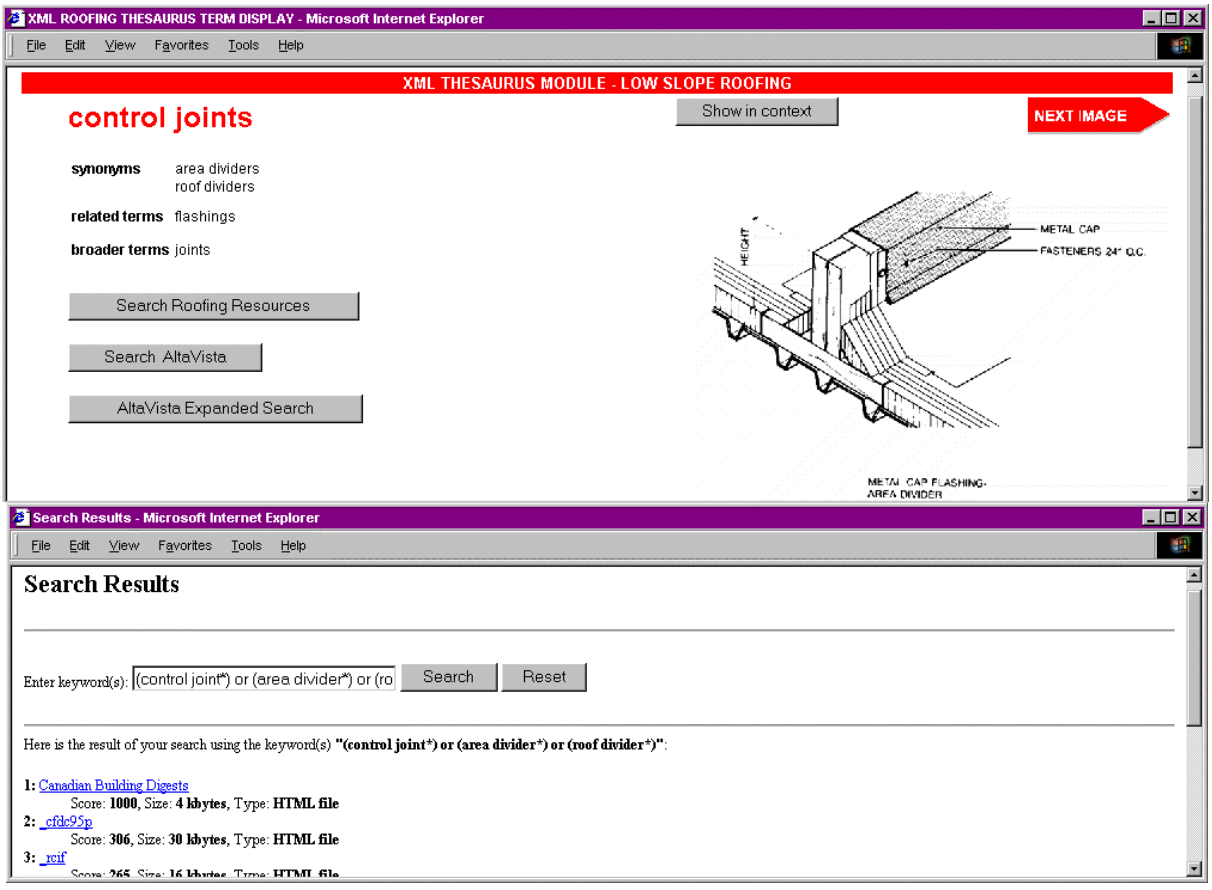

Figure 4. Screen shot of the XML prototype

Experiences with the prototype indicated some areas where further research might be needed. It would be useful to develop a classification of concepts in terms of their visualizability (e.g. abstract, direct, codeable, contextual) and to include the new conceptType attribute into our XML schema. Such classification might simplify information management by defining

constraints, i.e. by specifying which paths within the semantic network should be traversed for retrieving relevant materials for a particular request. Another point worth noting is that the images used in the prototype as well as concepts in the thesaurus represent exclusively classes. Such an application might be used for educational purposes, e.g. a visual dictionary or a library of construction methods. As in the AEC/FM practice information management involves representations of instances as well, mechanisms for switching between these two levels of abstraction are also required.

5. CONCLUSIONS

Thesaurus services can be used to complement the modelling technology by extending mechanisms for achieving semantic interoperability to the level of human-understandable information representations. We suggest the development of XML thesaurus modules for specific AEC/FM sub-domains as the need arises, which would be interoperable through mappings to the core domain schema and publicly available in a repository. Such modules would be used in server-based software components providing thesaurus services through their APIs to a wide variety of applications over the Internet.

The prototypes demonstrated the capability of thesaurus modules to enhance access to various types of unstructured information—from remote and unfamiliar Web pages to bitmap images from a local, controlled collection. The enhancement involves the capability to retrieve:

- documents that are highly relevant to a search although they do not contain the user's search terms and would therefore not be retrieved by a typical search service
- relevant images that are not indexed with the search term but are still accessible using knowledge captured in the conceptual network and code.

The enhancement is made possible by bringing together conceptual relationships and alternative human-understandable representations of concepts (i.e. alternative terms, text and images, alternative images).

It can reasonably be expected that thesaurus modules, especially if based on a standard domain schema, can also be used for:

- Accessing structured information (e.g. from XML or STEP physical files)
- Interrelating structured and unstructured information
- Inter-schema mappings and,
- Combined with more complex algorithms, tools and standards, for a variety of other purposes.

The Extensible Markup Language seems to be a particularly suitable tool for thesauri as it:

- Enables a variety of displays
- Simplifies design of thesaurus-based applications
- Makes thesauri available for use by remote applications over the Internet.

Serious applications, however, should wait until the sufficient maturity of infrastructure (e.g. browsers, search engines, XML and satellite standards) is achieved.

The work described in this paper touches only the tip of an iceberg. Our immediate further research will focus on transparent implementations of thesaurus services in real-life applications and environments, e.g. using a 3D model of a building to access different types of information interrelated through a thesaurus service invisible to the user. Many other topics briefly discussed in this paper remain open for further research:

- Co-ordination of vocabulary control in the domain

- Application of thesaurus techniques and experiences with inter-thesaurus mapping for the development, use, and mapping of domain schemas
- Development of applications using thesaurus services, integration of thesaurus modules with tools and standards aimed at integrating heterogeneous data sources.

6. REFERENCES

aecXML.org. (1999). aecXML Preliminary Specification: Working draft 0.87. Available from: <http://www.aecxml.org/index2.htm> Accessed [14 March 2000].

Buckland, M. & Plaunt, C. (1994). On the Construction of Selection Systems*, Library Hi Tech* 12(4), pp.15-28.

CERES/NBII Thesaurus Partnership Project Available from: <http://ceres.ca.gov/thesaurus/> Accessed [14 March 2000].

Chang S., Smith J., Meng H., Wang H. & Zhong D. (1997). Finding Image/Video in Large Archives: Columbia's Content-Based Visual Query Project, *D-Lib Magazine* February. Available from: <http://www.dlib.org/dlib/february97/columbia/02chang.html> [Accessed 14 March 2000].

Eastman C. (1999). Information exchange architectures for building models*. Durability of Building Materials and Components 8*, M. Lacasse & D. Vanier, eds, Institute for Research in Construction, Ottawa ON, Canada, pp. 2139-2156.

Foundation for Documents of Architecture, Architectural Drawing Advisory Group (1999). *Guide to the Description of Architectural Drawings.* Available from: <http://www.getty.edu/gri/standard/fda/index.htm> [Accessed 14 March 2000].

Froese, T. & Yu, K, (2000). "Architecture Issues For Distributed AEC/FM Systems", Submitted to *The 8th International Conference on Computing in Civil and Building Engineering, Stanford University, California, USA, August 14 – 17, 2000*.

Furnas, G. Landauer, T. Gomez, L. & Dumais, S. (1987). The vocabulary problem in human-system communication*. Communications of the Association for Computing Machinery*, 30(11), pp.964-971.

Industry Alliance for Interoperability. (1999). IFC Object Model, Release 2.0, [CD-ROM].

ISO (1999). ISO TC184/SC4/WG11 N088. EXPRESS-X Language Reference Manual: Working Draft, October 15, 1999.

Jarvelin, K. et al. (1996). A Deductive Data Model for Query Expansion, *SIGIR 96: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 18-22, 1996, Zurich, Switzerland*, eds Frei H. Et al., Konstanz: Hartung-Gorre, pp. 235-243.

Kosovac, B (1998). *Internet/Intranet and Thesauri*, Canadian Institute for Scientific and Technical Information, Internal Report, National Research Council Canada, Ottawa, Canada. Available from: <http://www.nrc.ca/irc/thesaurus/roofing/report_b.html> Accessed [14 March 2000].

Kosovac, B. Vanier, D. & Froese, T. (2000). Use of Keyphrase Extraction Software for Creation of an AEC/FM Thesaurus, *Electronic Journal of Information Technology in Construction* 5 pp.25-36. Available from: <http://itcon.org/2000/2/> Accessed [14 March 2000].

Lee, M. Baillie, S. & Dell'Oro, J. (1999). TML: A Thesaural Markup Language, *Proceedings of the 4th Australiasian Document Computing Symposium, Coffs Harbour, Australia, December 3, 1999*. Available from:

Library of Congress (1999). *Thesaurus for Graphic Materials I & II.* Available from: <http://lcweb.loc.gov/rr/print/tgm1/toc.html> [Accessed 14 March 2000].

National Information Standards Organization (1993). ANSI/NISO Z.39.19-1993. Guidelines for the Construction, Format, and Management of Monolingual Thesauri, Bethesda, MD: NISO Press.

National Library of Medicine. (1999). Unified Medical Language System (UMLS) Available from: <http://www.nlm.nih.gov/research/umls/> Accessed [14 March 2000].

Preaud, M. & Rio, M. (1978). Images sans histoire: Methode de description des images et classement informatique," in *First International Conference on Automation Processing of Art History Data and Documents, Pisa, 4-7 September 1978: Conference Transactions,* Barocche P. Et al., eds., vol.II, pp.248-257.

Shatford S. (1986). Analyzing the Subject of a Picture: A Theoretical Approach, *Cataloging & Classification Quarterly* 6, spring, pp.39-61.

Taylor M. (1999). Zthes: a Z39.50 Profile for Thesaurus Navigation. Available from: <http://www.n-four.demon.co.uk/mirk/zthes-02.html> [Accessed 14 March 2000].

Vanier, D.J. (1991) "A parsimonious classification system to extract project-specific building codes", *Computers and Building Regulations*, VTT Symposium, Vol. 125. Espoo, Finland, pp. 134-145.

Vanier, D.J. (1994). "Canadian thesaurus of construction science and technology: A Hypercard stack" in *Proceedings of the Joint CIB Workshops on Computers and Information in Construction*, Montreal, Que., Canada, CIB Proceedings, Vol. 165, eds. D.J. Vanier, J.R. Thomas, pp. 559-564.

Verhoef, M. Liebich, T. & Amor, R. (1995). "A Multi-Paradigm Mapping Method Survey" in *Modeling of Buildings Through Their Lifecycle: Proceedings, CIB Publication 180* Palo Alto, CA, USA, eds M Fischer, K Law, B Luiten, pp. 233-247

W3C (1998). Extensible Markup Language (XML) 1.0: W3C Recommendation 10-February-1998. Available from: <http://www.w3.org/TR/1998/REC-xml-19980210> Accessed [14 March 2000].

W3C (1999). XML Linking Language (XLink): World Wide Web Consortium Working Draft 26 July 1999. Available from: <http://www.w3.org/1999/07/WD-xlink-19990726> Accessed [14 March 2000].

W3C (2000). XML Query Requirements: W3C Working Draft 31 January 2000. Available from: <http://www.w3.org/TR/2000/WD-xmlquery-req-20000131> Accessed [14 March 2000].